Brand Extension Strategy Planning: Empirical Estimation of Brand–Category Personality Fit and Atypicality

RAJEEV BATRA, PETER LENK, and MICHEL WEDEL

Web Appendix A: Brand and Category Personality Model

The model specification for Equations (1) to (3) can be summarized as:

$$\begin{aligned}
&U_{i,j,b|c} = \varphi_i + \mu_j + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) + \varepsilon_{i,j,b|c} \\
&\mathrm{Var}(\varphi_i) = \tau^2 \text{ and } \mathrm{Var}\left(\varepsilon_{i,j,b|c}\right) = \sigma^2_{j,b|c} \\
&\mathrm{Var}(\alpha_i) = \mathrm{diag}\left(\lambda_f^2\right)\text{; } \mathrm{Var}\left(\alpha_{i,c}\right) = \mathrm{diag}\left(\lambda_{f,c}^2\right)\text{; and } \mathrm{Var}\left(\alpha_{i,b|c}\right) = \mathrm{diag}\left(\lambda_{f,b|c}^2\right)
\end{aligned} \tag{WA.1}$$

for subject i, personality item j, and brand b nested within category c. All of the random components are normally distributed with mean zero. These random components induce a covariance that reflects the structure of the data. Table WA.1 details the covariances. In our application, there are J=15 brand personality items, B=10 brands per category, C=3 categories, and P=5 brand personality dimensions. The total number of personality items are J×B×C = 450. An unconstrained covariance matrix has 101,475 unique parameters. The factor model has 666 parameters to model the covariance, excluding the J=15 mean parameters. There is one variance for the subject, random effect; 450 (J×B×C) error variances, 5 (P) variances for subject personality factors, 15 (C×P) variances for category personality factors, and 150 (B×C×P) variances for brand personality factors. The total number of loadings is 45. To identify the CFA model, we set the first, subject-level factor variance λ_1^2 to a constant. In special cases, such as C=1 or B=1, other constraints would have to be imposed on the model.

Table WA.1. Implied Covariance Structure from Hierarchical Factor Model

Items	Brands	Categories	Variance or Covariance
j	b	c	$Var(U_{i,j,b\|c}) = \tau^2 + \sum_{f=1}^{P}(\lambda_f^2 + \lambda_{f,c}^2 + \lambda_{f,b\|c}^2) l_{f,j,c}^2 + \sigma_{j,b\|c}^2$
j ≠ j*	b	c	$Cov(U_{i,j,b\|c}, U_{i,j^*,b\|c}) = \tau^2 + \sum_{f=1}^{P}(\lambda_f^2 + \lambda_{f,c}^2 + \lambda_{f,b\|c}^2) l_{f,j,c}\, l_{f,j^*,c}$
j & j*	b ≠ b*	c	$Cov(U_{i,j,b\|c}, U_{i,j^*,b^*\|c}) = \tau^2 + \sum_{f=1}^{P}(\lambda_f^2 + \lambda_{f,c}^2) l_{f,j,c}\, l_{f,j^*,c}$
j & j*	b ≠ b*	c ≠ c*	$Cov(U_{i,j,b\|c}, U_{i,j^*,b^*\|c^*}) = \tau^2 + \sum_{f=1}^{P}\lambda_f^2\, l_{f,j,c}\, l_{f,j^*,c^*}$

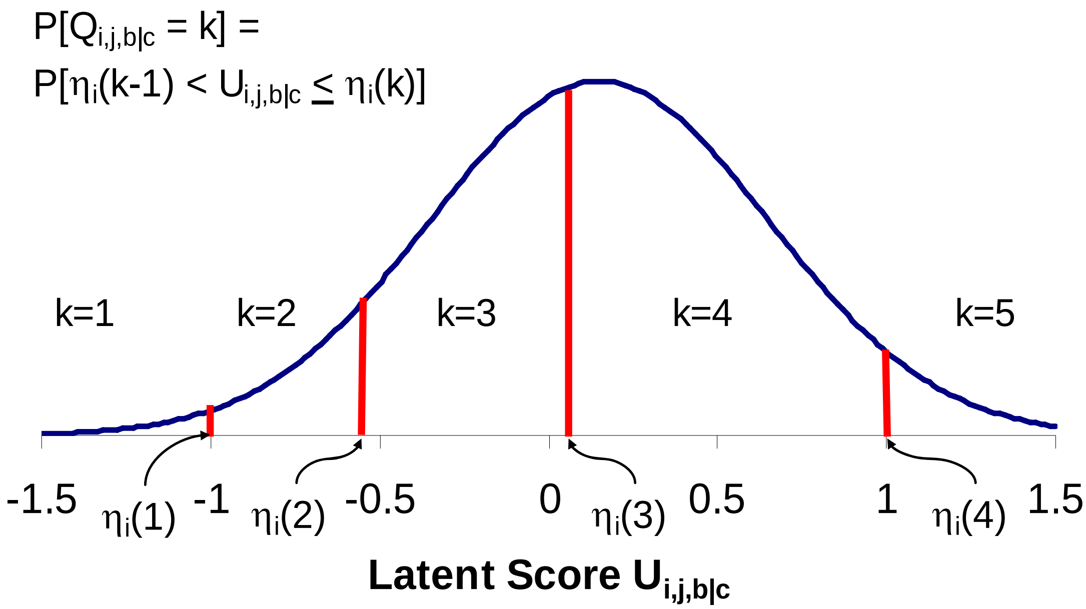
A cut-point model relates the latent variables $U_{i,j,b|c}$ to the observed ratings $Q_{i,j,b|c}$. All personality items were assessed on a nine-point scale. The cut-point model relates the ordinal personality scale responses to an underlying, normally distributed random variable. These models have a long tradition in statistics. Aitchison and Silvey (1957) proposed an ordinal probit model, and McCullagh (1980) used threshold models for ordinal regression. These models effectively relax the assumption that the observed variables have a normal distribution. The ordinal data can be very non-normal – e.g. skewed or bimodal – and the latent variable, which is used in the factor models, is still normally distributed. Gelfand, Smith, and Lee (1992) developed MCMC methods for ordered data. Rossi, Gilula, and Allenby (2001) and Johnson (2003) used heterogeneous threshold models to correct for scale-usage effects. The cut-point model assumes that respondent *i* selects scale point *k* for item *j* if the latent variable $U_{i,j,b|c}$ falls between two consecutive cut-points:

(WA.2) $$Q_{i,j,b|c} = k \quad \text{if and only if } \eta_i(k-1) < U_{i,j,b|c} \le \eta_i(k),$$

where there are K scale categories. Figure WA.1 shows the cut-points for the situation where $U_{i,j,b|c}$ has a normal distribution, the standard case, and there are 5 categories. In this paper, the cut-points are specific to the respondent and are ordered as follows:

$$\begin{aligned} &\eta_i(0) = -\infty \text{ and } \eta_i(1) = -1 \\ &\eta_i(k-1) < \eta_i(k) \text{ for } k = 2,\ldots,K-1. \\ &\eta_i(K-1) = 1 \text{ and } \eta_i(K) = \infty \end{aligned} \quad \text{(WA.3)}$$

To identify the model, we set $\eta_i(1)$ and $\eta_i(K-1)$ to –1 and 1, respectively. There are K-3 unknown cut-points per respondent, and these are estimated from the data. The model accounts for individual differences in the cut-points to accommodate idiosyncratic scale use, but we are not interested in the values of the cut-points for individual respondents per se.



The distribution of the observed ordinal personality data is derived from the latent variable and the cut-points are as follows:

$$\begin{aligned} P\left[Q_{i,j} = k\right] &= P\left[\eta_i(k-1) < U_{i,j} < \eta_i(k)\right] \text{ for } k = 1,\ldots,K \\ &= \int_{\eta_i(k-1)}^{\eta_i(k)} dF_{i,j}(u) \end{aligned}, \quad \text{(WA.4)}$$

where $F_{i,j}$ is the cumulative distribution of $U_{i,j,b|c}$, which is taken to be the normal distribution in this paper.

We use standard prior distributions. The prior distributions for all means $\{\mu_j\}$ are normal with mean zero and large variances. In CFA, the free loadings are assumed to be positive, and their prior distribution is a normal distribution with mean zero that is truncated below at zero, also known as “half-normal.” All variance parameters come from an inverse gamma distribution. In the cut-point model, the free parameters are uniformly distributed on $\eta_i(2) < \ldots < \eta_i(K-2)$.

A final feature of our approach is that it enables imputation of missing data (if any), arising due to the split-structure of the questionnaire. Splitting the questionnaire, by randomly assigning blocks of questions to subjects as we did in the present study, creates missing data, hindering the application of traditional factor models. We impute the missing data at the same time as we estimate the model, by drawing from the “predictive distribution” of the data, an approach that is made possible by the application of the Gibbs sampler (Gelfand and Smith 1990). The estimation of the model adapts the MCMC procedures described in Ansari and Jedidi (2000), Ansari Jedidi and Dube (2002), and Song and Lee (2001 and 2002).

Web Appendix B: Summary of MCMC Sampling Scheme

We assume that the reader is familiar with the updating the full conditional distribution for standard models using MCMC.

Update cut-points $\{\eta_i(k)\}$ for k = 2, …, K-2.

The ordinal response $Q_{i,j,b|c} = k$ if $\eta_i(k-1) < U_{i,j,b|c} \leq \eta_i(k)$. Thus, the full conditional distribution for $\eta_i(k)$ is uniform on the interval with lower endpoint $\max(U_{i,j,b|c} = k, \eta_i(k-1))$ and upper endpoint $\min(U_{i,j,b|c} = k+1, \eta_i(k+1))$. We need to include $\eta_i(k-1)$ and $\eta_i(k+1)$ in the updating if none of the observations for subject i take the values k or k+1, respectively. We exclude missing observations.

Update the latent variables $\{U_{i,j,b|c}\}$.

If the ordinal response $Q_{i,j,b|c} = k$, then $\eta_i(k-1) < U_{i,j,b|c} \leq \eta_i(k)$ where the latent variables $\{U_{i,j,b|c}\}$ are independently distributed normal distributions with means and variances:

$$E\left[U_{i,j,b|c}\right] = \phi_i + \mu_j + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) \text{and } \mathrm{Var}\left(U_{i,j,b|c}\right) = \sigma^2_{j,b|c}.$$

If $Q_{i,j,b|c} = k$ is observed, then generate $U_{i,j,b|c}$ from a truncated normal distribution on the interval $\eta_i(k-1)$ to $\eta_i(k)$ with the above mean and variance. We use the inverse cdf normal method for generating truncated normal distributions. If g and G are the unconstrained normal density and cumulative distribution functions for U, then generate $a < U < b$ by: $U = G^{-1}\{ V[G(b) - G(a)] + G(a) \}$ where V is a uniform random deviate on (0,1). If $Q_{i,j,b|c}$ is unobserved, then generate $U_{i,j,b|c}$ from a unconstrained normal distribution with the above mean and variance.

We then analyzed the rest of the model with one of two methods: using the generated $\{U_{i,j,b|c}\}$ for both the observed and missing data, and using just the observed data. The first method corresponds to Bayesian data imputation methods. The results of the two methods are theoretically identical and the numerical results were similar. The paper reports the results using just the observed $\{U_{i,j,b|c}\}$ because they are missing by the design of the survey: subsets of subjects evaluated different subsets of the 30 brands. If there were non-responses to items on the questionnaire, we would impute the missing values. However, our dataset had compete responses to all items on each questionnaire.

Update the subject-level random effect $\{\varphi_i\}$.

Because the random effects have $N(0,\tau^2)$ distribution, and

$$\phi_i = U_{i,j,b|c} - \left\lfloor \mu_j + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) + \varepsilon_{i,j,b|c} \right\rfloor$$

has a normal likelihood, the full conditional is also normally distributed with the standard, updated mean and variance.

Update the variance τ^2 of the random effects.

Because the prior distribution for τ^2 is an inverse Gamma distribution, and the $\{\varphi_i\}$are iid $N(0,\tau^2)$, the full conditional is also inverse Gamma with the standard updating of the prior parameters.

Update the population mean μ_j for item j.

Because the prior distribution for μ_j is a normal distribution and

$$\mu_j = U_{i,j,b|c} - \left\lfloor \phi_i + L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) + \varepsilon_{i,j,b|c} \right\rfloor$$

has a normal likelihood, its full conditional is also normally distributed with the standard, updated mean and variance.

Update the error variance terms $\{\sigma_{j,b|c}^2\}$.

Because the prior distribution of $\sigma_{j,b|c}^2$ is inverse Gamma distribution and the $\{U_{i,j,b|c}\}$ has a normal distribution with variance $\sigma_{j,b|c}^2$, the full conditional of $\sigma_{j,b|c}^2$ has an inverse Gamma distribution with the standard updating of the prior parameters.

Update the factor loadings $\{L_c\}$.

In CFA, the priors of the free loadings have truncated normal distributions that are greater than 0. Because the loadings in

$$L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) = U_{i,j,b|c} - \left\lfloor \phi_i + \mu_j + \varepsilon_{i,j,b|c} \right\rfloor$$

given the factor scores have a normal likelihood where the variance $\sigma_{j,b|c}^2$ are scaled by the factor scores, the full conditional also has a positive, truncated normal distribution with the standard updating of the mean and variances. We use the inverse normal cdf method for generating truncated normal deviates. In EFA, the priors for the free loadings are normally distributed, and the full conditionals also have normal distributions with the standard updating of parameters.

Update the factor scores $\{\alpha_i, \alpha_{i,c}, \alpha_{i,b|c}\}$.

The model assumes that the factor scores are iid normals with mean 0 and variances $\{\text{diag}(\lambda_f^2), \text{diag}(\lambda_{f,c}^2), \text{diag}(\lambda_{f,b|c}^2)\}$. Because the factors scores in

$$L'_{j,c}\left(\alpha_i + \alpha_{i,c} + \alpha_{i,b|c}\right) = U_{i,j,b|c} - \left[\phi_i + \mu_j + \varepsilon_{i,j,b|c}\right]$$

given the loadings have a normal likelihood with scaled variances, their full conditional distribution also is a normal distribution with the standard updating of the means and variances. In the updating one needs to keep track of which observations are used for which factors. All of the observations for subject i are used in updating $\alpha_{i.}$ All of the observations for subject i and category c are used in updating $\alpha_{i,c.}$ Only the observations for subject i and brand b in category c is used in updating $\alpha_{i,b|c}$.

Update the factor variances $\{\lambda_f^2, \lambda_{f,c}^2, \lambda_{f,b|c}^2\}$.

Because the prior distributions for these variances are inverse Gamma and the factor scores are normally distributed with mean 0 and these variances, the full conditionals are also inverse Gamma distributions with the standard updating. To identify the model in CFA, the first component of λ_f^2 is set to a constant and is not updated.

We fitted the model with Markov chain Monte Carlo (MCMC) methods (Gelfand and Smith 1990), using a program we wrote in GAUSS. The MCMC algorithm ran for 100,000 iterations. The initial transition period consisted of 50,000 iterations, which were not used in estimation. Of the next 50,000 iterations, every tenth iterate was used in the analysis for a total of 10,000. Traces of the iterations indicate that the chain reached the stable distribution well before the 50,000 iteration, and simulation studies using artificial data indicated that 30,000 iterations were more than sufficient: the chains using simulated data frequently converged within 3000 iterations and remained stable at the true parameters thereafter. The algorithm simultaneously imputed the data for brands that a respondent did not evaluate, thus simplifying the analysis.